

Enhancing 2D Pixels to 3D Models Generation Via Input Normalization



Hsin-Fei Wu, Chieh-Ting Hsu, Pei-Ti Shi, Chang-Jui Lin

Introduction

Background

- 2D images preserve memories but remain visually flat and limited for 3D creation
- Converting single 2D images into 3D expands possibilities in animation, gaming, and digital content

Core challenges

- Background noise interferes with model clarity
- Low resolution blurs important details
- Overlapping poses cause ambiguity
- Viewpoint distortion alters shape perception

Our idea

We propose an input normalization pipeline that improves the quality of the source image before 3D generation. By standardizing the input, we aim to produce more accurate, stable, and usable 3D outputs.

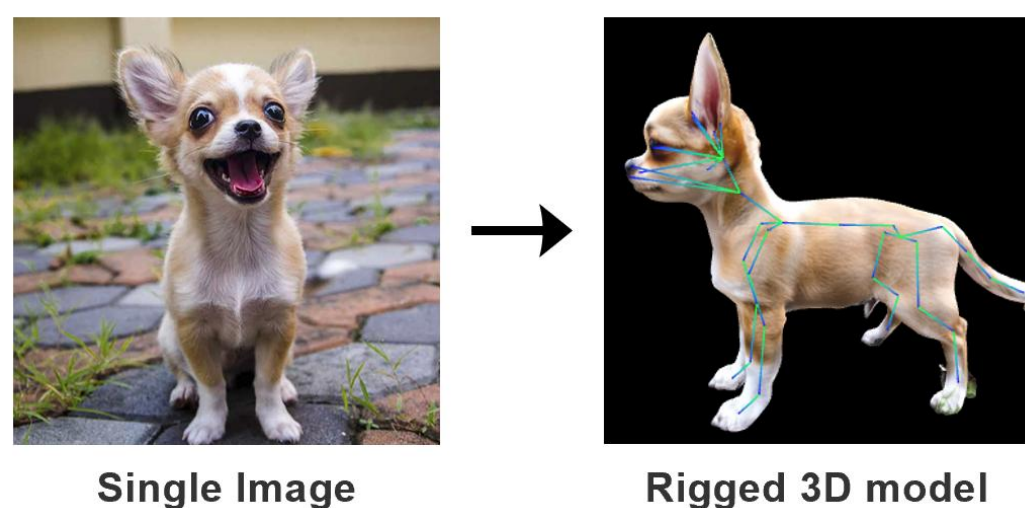


Figure 1 — Stable Rigged 3D Generation from a Single Image

Method

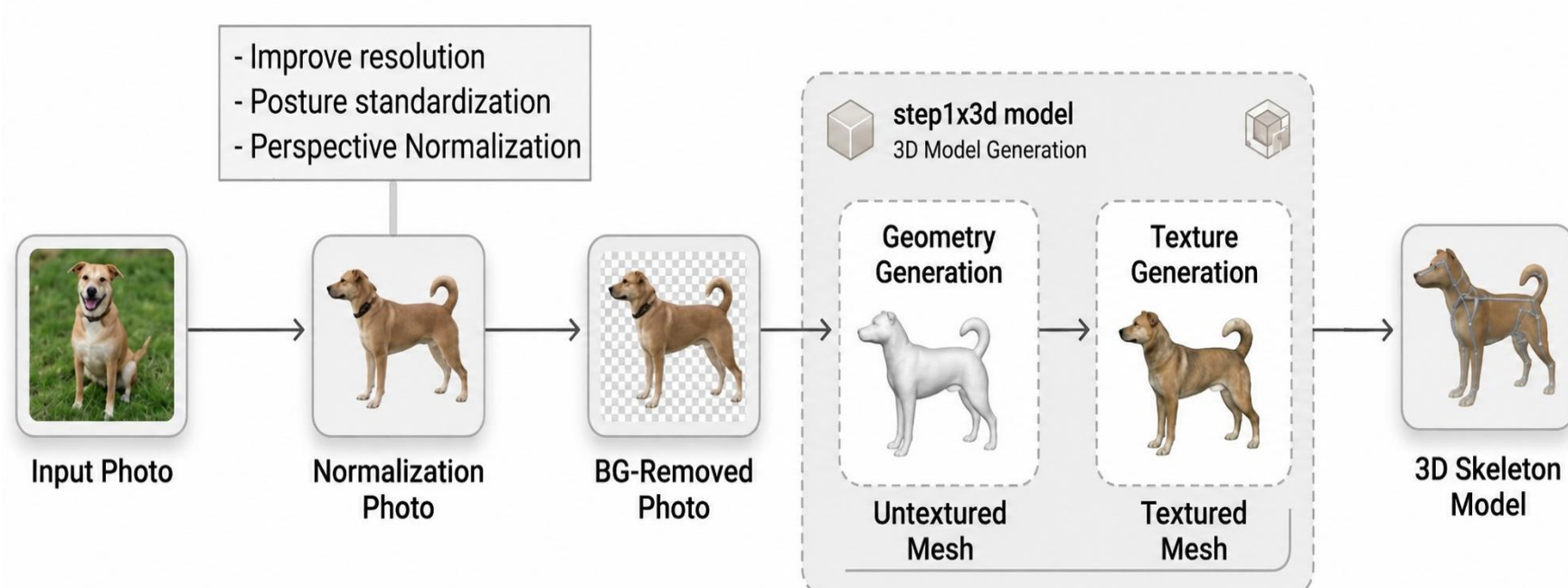


Figure 2 — System pipeline

Stage 1: Input Normalization

1. Pose normalization
2. Viewpoint normalization
3. Resolution enhancement
4. Background removal (Gemini 2.5 flash Image Model and BiRefNet)

Stage 2: 3D Reconstruction

1. Geometry generation
2. Texture generation (Step 1X-3D Model)

Stage 3: Rigging & Applications

1. Skeleton generation
2. Animation
3. Voxelization

Problem and Solution

Problem	Cause	Impact	Solution
Background Noise	Background Interference	Incorrect Texture Generation	Background Removal
Low Resolution	Insufficient Visual Details	Blurred Texture Reconstruction	Resolution Enhancement
Pose Ambiguity	Overlapping Limbs / Unusual Poses	Geometry & Skeleton Prediction Errors	Pose Normalization
Viewpoint Distortion	Missing Depth Information	Shape Deformation	Viewpoint Normalization

Results

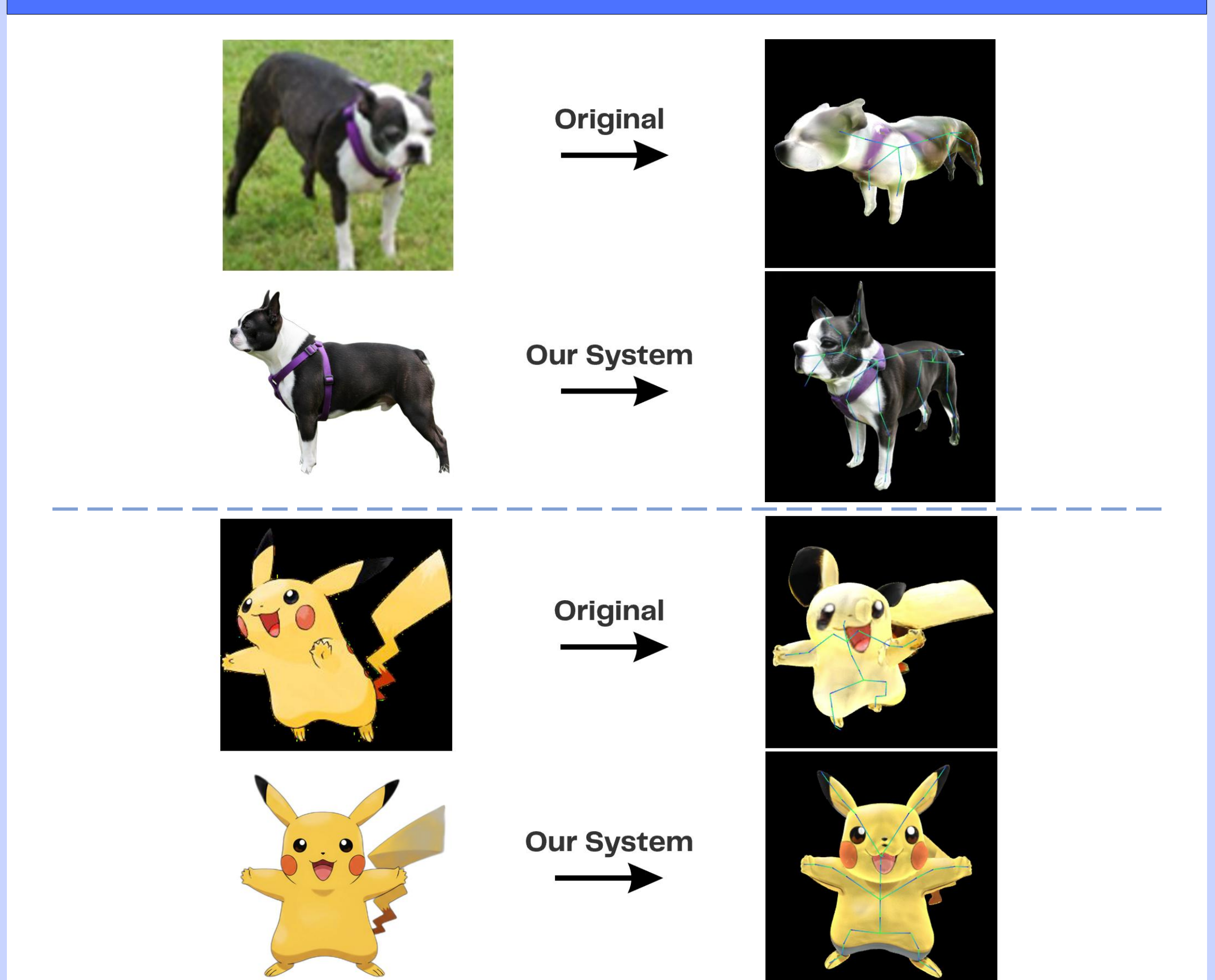


Figure 3, Figure, 4 — Comparison between original input and normalized input in 3D generation

Conclusion

- Analyzed key failure factors in single-image 3D reconstruction, including background interference, pose ambiguity, and viewpoint distortion.
- Proposed an input normalization pipeline to improve geometry reconstruction, texture generation, and skeleton prediction quality.
- Developed an end-to-end automated system that converts a single 2D image into rigged 3D models and voxelized outputs.
- Enhanced the stability and practical usability of single-image-to-3D generation workflows.